Chapter 6

Data Set/Data Set Collection Contents and Naming

One of the objectives of the PDS is to introduce consistency in the contents, organization and naming of planetary data sets. The PDS has introduced the concept that an archive quality data set collection or data set must include everything that is needed to understand and utilize the data. Towards this goal, the PDS has worked with the PDS Discipline Nodes, numerous Flight Projects, individual scientists and programmers, and the NSSDC to develop approaches to ensure that this consistency is achieved.

Figure 6.1 shows the relationships between Data Set Collection, Data Sets, and Data Products. Figure 6.2 shows the logical and physical relationships.
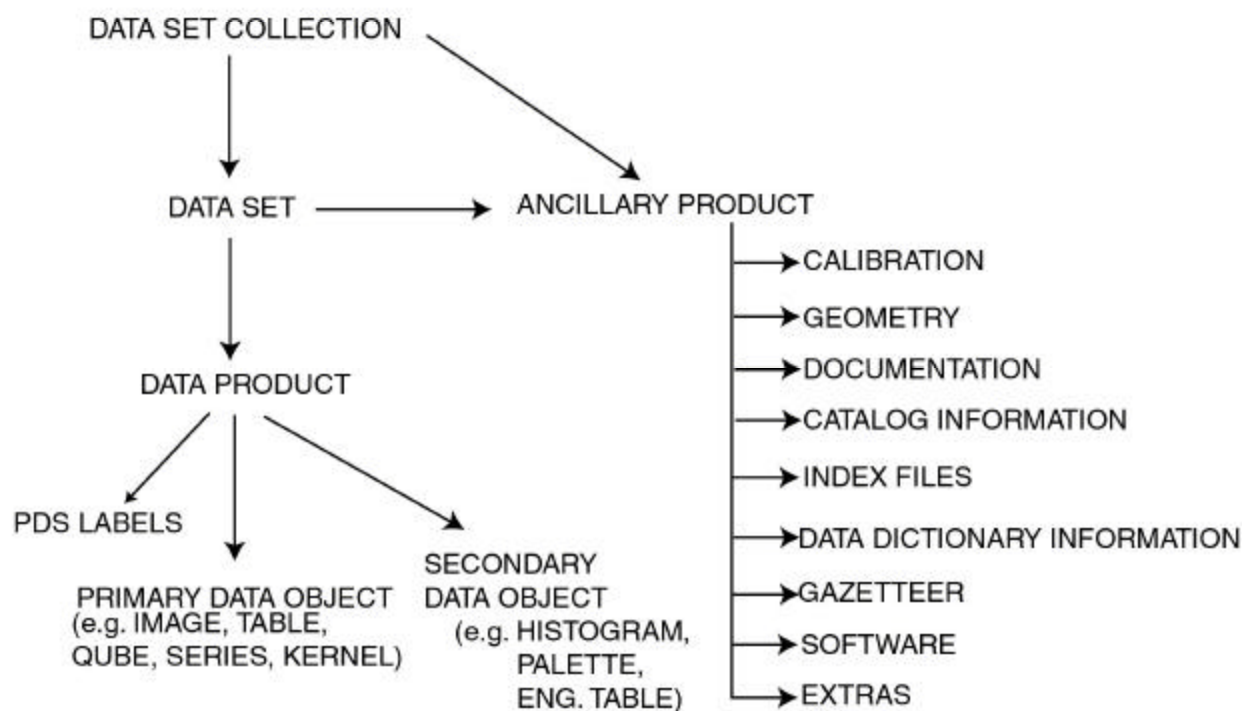
CONTENTS

Figure 6.1 Data Set Collection, Data Sets, and Data Products
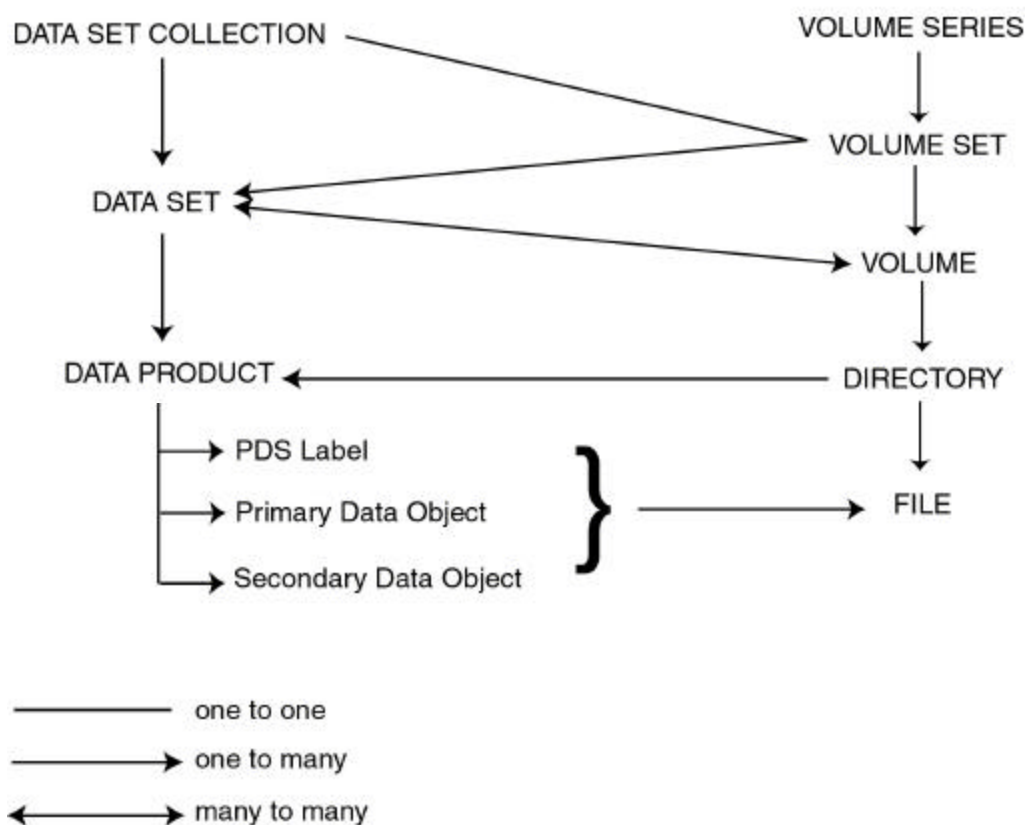
LOGICAL/PHYSICAL RELATIONSHIPS

Figure 6.2 Logical and Physical Relationships of Data Products

6.1   Data Set/Data Set Collection Contents

Data Set Collection and Data Set defined:

Data Set Collection - A data set collection consists of data sets that are related by observation type, discipline, target, or time, and therefore are to be treated as a unit, to be archived and distributed as a group (set) for a specific scientific objective and analysis. One of the primary considerations in creating a data set collection is that the collection as a whole provides more utility than any individual data set within the collection. Further, the individual data set(s) may be ineffective unless used in concert with the other data sets in the collection.

Data Set - The accumulation of data products, secondary data, software, and documentation, that completely document and support the use of those data products. A data set can be part of a data set collection.

A data set collection or a data set may include all of the following:

Data Product - A labeled grouping of data resulting from a scientific observation. Examples of data products include planetary images, spectrum tables, and time series tables. A data product is a component of a data set.

Calibration data - Calibration files used in the processing of the raw data or needed to use the data.

Geometry data - Relevant files (e.g., SEDRs, SPICE kernels) needed to describe the observation geometry.

Documentation - All the textual material which describes the mission, spacecraft, instrument, and data set. This can include references to science papers, or the actual papers.

Catalog Information - High-level descriptive information about a data set (e.g. mission description, spacecraft description, instrument description), expressed in Object Description Language (ODL) which is suitable for loading into a catalog.

Index Files - Information which allows the user to locate the data of interest, such as a table mapping latitude/longitude ranges to file names.

Data Dictionary Information - A portable version of the *Planetary Science Data Dictionary* which is pertinent to the data set. The dictionary is expressed in ODL.

Gazetteer - Information about the named features on a target body associated with the data sets.

Software- The software libraries, utilities, or application programs to access/process the data objects.

All data sets submitted to the PDS shall include the software used and/or algorithms for original data reduction, processing, calibration and, decalibration of the data, or documentation stating how to obtain such software. When software accompanies a data set, the source code, build instructions, and software documentation shall be included.

There are several other types of data set software which may be provided with a data set:

1.   Special software which is developed and maintained for certain hardware platforms. This is often a refined version of the processing software developed for mission data analysis.

2.   Utilities which allow a user to select parameters from the data set and to extract these parameter values to a data file based on certain key values (event time, for example). The output format should be a simple ASCII table or one of the other generic PDS data object formats. This is a minimum level of access for conducting a peer review of a data set.

3.   Data analysis tools such as plotting programs.

6.2   Data Set Naming and Identification

This standard contains instructions for naming a PDS data set and forming a Data Set Identifier. Every PDS data set shall be given a DATA_SET_NAME and DATA_SET_ID, both formed from seven components. All components are required except for the Data Set Type and Description components. These components are described in section 6.4.

The only characters allowed within a data_set_id are the upper case alphanumeric set (A-Z, 0-9), a forward slash (/), a period (.), and a hyphen (-). The period is only used with numerics, i.e., V1.0 or 12.5SEC. No other special characters are allowed (e.g., underscore (_)).

Multiple instrument hosts, instruments, or targets shall be referenced in a DATA_SET_NAME or DATA_SET_ID by concatenation of the values with a forward slash (/) which is interpreted as "and."

The data set identifier (DATA_SET_ID) shall not exceed 40 characters in length. Each component shall be the acronym rather than a full length name used in forming the DATA_SET_NAME. Within the data_set_id, acronyms shall be separated by hyphens.(See section 6.4 for valid acronyms.)

A DATA_SET_NAME shall not exceed 60 characters in length. Where the character limitation is not exceeded, the full length name of each component should be used. If the full length name is too long, an acronym shall be used to abbreviate components of the name. (See section 6.4 for valid full length names and acronyms.)

The intent of the data set name (DATA_SET_NAME) and identifier (DATA_SET_ID) is primarily to uniquely identify the data set.

The components of the DATA_SET_NAME and DATA_SET_ID are:

Instrument host
Target
Instrument
Data processing level number
Data set type (optional)
Description (optional)
Version number

Example:

• Full length data set name: Mariner 9 and Viking Orbiter 1 and Viking Orbiter 2 Mars Imaging Science Subsystem and Visual Imaging Subsystem derived cloud data Version 1.0

• DATA_SET_NAME = "MR9/V01/V02 MARS IMAGING SCIENCE SUBSYSTEM/VIS 5 CLOUD V1.0"

• DATA_SET_ID = "MR9/V01/V02-M-ISS/VIS-5-CLOUD-V1.0"

In this example,   Instrument hosts are Mariner 9, Viking Orbiter 1 and Viking Orbiter 2
                     Target is Mars
                     Instruments are the Imaging Science Subsystem and Visual Imaging Subsystem
                     Data Processing Level number is 5
                     Description is CLOUD
                     Version number is V1.0
                     The optional Data set type is not used in this example.

6.3   Data Set Collection Naming and Identification

This standard contains instructions for naming a PDS data set collection and forming an identifier. A data set collection consists of data sets that are related by observation type, discipline, target, or time (which are treated as a unit), for a specific scientific purpose.

A data set collection will contain data sets that may cover several targets, be of different processing levels, and have different instrument hosts and instruments. Since the individual data sets will be identified by their own data set names, some of this information is not necessary to repeat at the collection level. Therefore, the DATA_SET_COLLECTION_NAME uses a subset of the DATA_SET_NAME components in addition to a new component, collection name, which identifies the group of related data sets.

The DATA_SET_COLLECTION_NAME and DATA_SET_COLLECTION_ID are formed from the six components listed below. All are required, except for data processing level number, data set type, and description. However, it is recommended that data set type or description be used whenever possible.

The only characters allowed within a DATA_SET_COLLECTION_ID are the upper case alphanumeric set (A-Z, 0-9), a forward slash (/), a period (.), and a hyphen (-). The period is only used with numerics, i.e., V1.0 or 12.5SEC. No other special characters are allowed (e.g., underscore (_)).

Multiple targets or data processing levels shall be referenced in the data set collection name or identifier by concatenation of the values with a forward slash (/) which is interpreted as "and."

A DATA_SET_COLLECTION_NAME shall not exceed 60 characters in length. Where the character limitation is not exceeded, the full length name of each component should be used. If the full length name is too long, an acronym shall be used to abbreviate it. (See Section 6.4 for valid full length names and acronyms.)

The DATA_SET_COLLECTION_ID shall not exceed 40 characters in length. Each component shall be the acronym rather than a full length name used in forming the DATA_SET_COLLECTION_NAME. Within the DATA_SET_COLLECTION_ID, acronyms shall be separated by hyphens. (See Section 6.4 for valid acronyms.)

The components of the DATA_SET_COLLECTION_NAME and DATA_SET_COLLECTION_ID are:

Collection name
Target
Data processing level number (optional)
Data set type (optional)
Description (optional)
Version number

Example:
The Pre-Magellan Data Set Collection contains radar and gravity data similar to the kinds of data that Magellan collected and was used for pre-Magellan analyses of Venus and for comparisons to actual Magellan data.

• Full-length data set collection name: Pre-Magellan Earth, Moon, Mercury, Mars, and Venus Resampled and Derived Radar and Gravity Data Version 1.0

• DATA_SET_COLLECTION_NAME = "PRE-MAGELLAN E/L/H/M/V 4/5 RADAR / GRAVITY DATA V1.0"

• DATA_SET_COLLECTION_ID = "PREMGN-E/L/H/M/V-4/5-RAD/GRAV-V1.0"

6.4   Description of Name and ID Components

If the information needed to describe your data is not listed, consult the PDS Data Engineer to determine what the appropriate acronyms are for you to use.

When a reference is made to the PSDD, see the standard values list for the data elements.

1. **Instrument host** component valid values are:
   full length names: INSTRUMENT_HOST_NAME data element in the PSDD
   acronyms: INSTRUMENT_HOST_ID data element in the PSDD
   exceptions: for Earth based data sets with no instrument host defined, the default
                                   value of EAR is recommended.

2. **Collection name** component valid values may be one of the following:

   GRSFE   Geological Remote Sensing Field Experiment
   IHW   International Halley Watch
   PREMGN   Pre-Magellan

3. **Target** component valid values are:

   full length names: TARGET_NAME data element in the PSDD
   acronyms: one of the following target IDs

Target ID	Target Name
A	Asteroid
C	Comet
CAL	Calibration
D	Dust
E	Earth
H	Mercury
J	Jupiter
L	Moon
M	Mars
MET	Meteorite
N	Neptune
P	Pluto
R	Ring
S	Saturn
SA	Satellite
SS	Solar System
U	Uranus
V	Venus
X	Other, ex. Checkout
Y	Sky

NOTE: Satellites or rings shall be referenced in a DATA_SET_NAME and DATA_SET_ID by the concatenation of the satellite or ring identifier with the associated planet identifier; for example:

JR = Jupiter's rings
JSA = Jupiter's satellites

If Jupiter data are also included in the ring and/or satellite data set, then only Jupiter, J, is referenced as the target.

In cases where there are data sets of comets or asteroids this component represents the TARGET_TYPE rather than the target name, for example:

A = Asteroid   CAL = Calibration
C = Comet   MET = Meteorite

Valid values for the TARGET_TYPE data element are found in the PSDD.

4. **Instrument** component valid values are:

full length names:	INSTRUMENT_NAME data element in the PSDD
acronyms:	INSTRUMENT_ID data element in the PSDD
exceptions:	ENG or ENGINEERING for engineering data sets
	SPICE for SPICE data sets
	GCM for Global Circulation Model data
	SEDR for supplemental EDR data
	POS for positional data

5. **Data processing level number**

This component is the National Research Council (NRC) Committee on Data Management and Computation (CODMAC) data processing level number.

Normally a data set contains data of one processing level. PDS recommends that data of different processing levels be treated as different data sets. However, if it is not possible to separate the data, then a single data set with multiple processing levels will be accepted. Use the following when specifying the data processing level number component of the data set identifier and name:

(a) the processing level number of the largest subset of data or
(b) the highest processing level number if there is no predominant subset.

DATA LEVEL NUMBER (CODMAC AND NASA LEVELS)

Level	Proc. Type	Data Processing Level Description
1	Raw Data	Telemetry data with data embedded.
2	Edited Data	Corrected for telemetry errors and split or decommutated into a data set for a given instrument. Sometimes called Experimental Data Record. Data are also tagged with time and location of acquisition. Corresponds to NASA Level 0 data.
3	Calibrated Data	Edited data that are still in units produced by instrument, but that have been corrected so that values are expressed in or are proportional to some physical unit such as radiance. No resampling, so edited data can be reconstructed. NASA Level 1A.
4	Resampled Data	Data that have been resampled in the time or space domains in such a way that the original edited data cannot be reconstructed. Could be calibrated in addition to being resampled. NASA Level lB.
5	Derived Data	Derived results, as maps, reports, graphics, etc. NASA Levels 2 through 5.
6	Ancillary Data	Nonscience data needed to generate calibrated or resampled data sets. Consists of instrument gains, offsets; pointing information for scan platforms, etc.
7	Correlative Data	Other science data needed to interpret spaceborne data sets. May include ground-based data observations such as soil type or ocean buoy measurements of wind drift.
8	User Description	Description of why the data were required, any peculiarities associated with the data sets, and enough documentation to allow secondary user to extract information from the data.
N	N	Not Applicable

6. **Data set type**

Normally, the data processing level (CODMAC) component is sufficient to be able to identify the type or level of data. However, if additional identification is desired, this component may be used. The following is a list of valid values (both full length names and acronyms) that may be used for this component.

NOTE: Several of the values in this table are currently unique to a particular mission (e.g. BIDR, MIDR were used on Magellan). These values should also be used on other missions, if deemed appropriate.

Acronym	Description
ADR	Analyzed Data Record
BIDR	Basic Image Data Record
CDR	Composite Data Record
CK	SPICE CK (Pointing Kernel)
DDR	Derived Data Record (possibly multiple instruments)
DIDR	Digitalized Image Data Record
DLC	Detailed Level Catalog
EDC	Existing Data Catalog
EDR	Experiment Data Record
EK	SPICE EK (Instrument Kernel)
GDR	Global Data Record
IDR	Intermediate Data Record
IK	SPICE IK (Instrument Kernel)
LSK	SPICE LSK (Leap Second Kernel)
MDR	Master Data Record
MIDR	Mosaicked Image Data Record
ODR	Original Data Record
PCK	SPICE PCK (Planetary Constants Kernel)
PGDR	Photograph Data Record
RDR	Reduced Data Record
REFDR	Reformatted Data Record
SDR	System Data Record

SEDR	Supplementary Experiment Data Record
SPK	SPICE SPK (Ephemeris Kernel)
SUMM	Summary (data) (to be used in the browse function)
SAMP	Sample data from a data set (not subsampled data)

7. **Description**

The following is a list of example values (both full length names and acronyms) that could be used for this component.

While the description is optional, it allows the user to provide information to help describe the data set, such as identifying a specific comet or asteroid.

Acronym	Description
ALT/RAD	Altimetry and Radiometry
BR	Browse
CLOUD	Cloud
ELE	Electron
ETA-AQUAR	Meteor Eta-Aquarius
FULL-RES	Full Resolution
GIACOBIN-ZIN	Comet Giacobini Zinner
HALLEY	Comet Halley
ION	Ion
LOS	Line of Sight Gravity
MOM	Moment
PAR	Parameter
SA	Spectrum Analyzer
SA-4.0SEC	Spectrum Analyzer 4.0 second
SA-48.0SEC	Spectrum Analyzer 48.0 second

8.   **Version number**

The rules for determining version numbers for PDS Data Sets/Data Set Collections are as follows:

(a)   If there is not a previous version of the PDS data set/data set collection, then use Version 1.0.

(b)   If a previous version exists, then consider the following:

  i.   If the data sets/data set collections contain the same set of data, but use a different medium (e.g., CD-ROM), then no new version number is required (i.e. no new data set identifier). The inventory system will handle the different media for the same data set.

  ii.   If the data sets/data set collections contain the same set of data, but have minor corrections or improvements such as a change in descriptive labeling, then the version number is incremented by a tenth. For example, V1.0 becomes V1.1.

  iii.   If a data set/data set collection has been reprocessed, using, for example, a new processing algorithm or different calibration data, then the version number is incremented by one (V1.0 would become V2.0). Also, if one data set/data set collection contains a subset, is a proper subset, or is a superset of another, then the version number is incremented by one.